Adaptive Classification Improves Control Performance in ERP-Based BCIs

Sven Dähne, Johannes Höhne, and Michael Tangermann

BBCI Lab, Machine Learning Dept., Berlin Institute of Technology, Berlin, Germany

Abstract

This contribution investigates the effects of applying an unsupervised adaptation mechanism to linear classifiers for Brain-Computer Interfaces (BCI). Specifically, we track changes in the first two moments of the unlabeled data distribution. Changes are adaptively compensated by recalculating the classifier based on short, consecutive data segments. The approach is validated on three auditory oddball data sets containing a total of N=37 subjects, of which 6 were used for model selection and the remaining 31 for validation. We find a significant performance increase (up to 14%) for the adaptive scheme compared to a fixed classifier. The increase is largest for subjects with low performance.

1 Introduction

Brain-Computer Interface (BCI) systems observe and analyze neural signals recorded for example via EEG. A BCI aims to translate these signals into control commands for a computer program or neural prosthesis. In order to reach this goal, BCIs must not only be accurate in their class-discrimination, they must also be robust with respect to the inherently low signal to noise ratio. Furthermore, the robustness to changes in the data distributions of ongoing or task-specific signals, often referred to as non-stationarities, is a desirable characteristic of a BCI.

In principle two approaches can be identified to reach the latter goal: finding signal representations that are maximally invariant with respect to such non-stationarities [1] or trying to detect and to compensate them [2]. Adaptive methods, which fall into class of the compensating approach have been successfully applied in the context of BCIs based on motor-imagery [3].

In this work we pursue a similar approach in the context of BCIs based on auditory event-related potentials (ERP). Specifically, we track changes in moments of the feature distributions over time and carefully adapt the classifier to alleviate the effects of non-stationarities.

2 Methods

2.1 Data

The EEG data sets used for offline-analysis stem from three separate auditory oddball ERP BCI studies, referred to as PASS2D, AMUSE, and WARP, respectively. All data sets were recorded using a 63 channel layout and 1000 Hz sampling rate. The data was band-pass filtered with a pass band of 0.4 to 40 Hz and down-sampled to 100 Hz prior to analysis. Epochs were extracted between -150 ms and 800 ms relative to the stimulus onset. Details specific to each of the studies are as follows [PASS2D, AMUSE, WARP]: number of stimulus epochs in calibration phase = [3402, 4320, 972], number of stimulus epochs in the online phase = [11987, 8100, 10692], target to non-target ratio = [1:8, 1:5, 1:8], number of participants = [10, 21, 6]. Further details about the PASS2D and AMUSE paradigms are provided by [4] and [5].
2.2 Classification

2.2.1 Feature Extraction

After the calibration phase, temporal features were extracted from the raw epochs and utilized to train a linear classifier. In this context, a temporal feature is defined as the mean EEG value of a post-stimulus interval. The intervals are the same for all channels and were chosen by hand by the experimenter (PASS2D and AMUSE) or by a heuristic (WARP). Depending on the subject, a number of 2, 3, or 4 intervals are chosen. This yields a feature vector of dimensionality $N_{fv}$, which is the product of the number of intervals and the number of recording channels.

2.2.2 Fixed Classifier

The data from the calibration measurements was used to train a Linear Discriminant Analysis (LDA) classifier. This classifier was not adapted during the (offline) classification of the online data and thereby represents the baseline for comparison. LDA was used, as this method has proven to yield robust and accurate results in many studies [6] and allows for inspection and analysis of the decision mechanism (i.e. inspection of the feature-weights). Thus, the classifier output for a data sample $x$ is given by

$$LDA_{fix} = w^T x + b,$$

with

$$w = C^{-1}(\mu_2 - \mu_1),$$

$$b = -0.5 w^T (\mu_1 + \mu_2).$$

In this formalism, $C^{-1}$ is the inverse of the covariance matrix of the pooled data and $\mu_1$ and $\mu_2$ denote the mean of the class target and non-target, respectively. Since $N_{fv}$ can become quite large (e.g. $N_{fv} = 252$ for 4 intervals and 63 EEG channels) relative to the number of training samples, we applied a regularization method (shrinkage [6]) to estimate $C$.

The binary decision about whether $x$ belongs to either of the two classes is based on the sign of $LDA_{fix}$. This classifier is termed fixed because it is computed on the calibration data and remained fix thereafter.

2.2.3 Adaptive Classifier

Note that the computation of $w$ (eq. 2) consists of two factors, namely $C^{-1}$ and the difference of the class means ($\mu_2 - \mu_1$). Only the second factor requires class labels which is only available during calibration. The first factor, on the other hand, can be computed in an unsupervised manner during calibration as well as during the online phase [2].

Thus, after being initialized on the calibration data, the classifier is made adaptive by updating the pooled mean and pooled covariance. These two moments are re-estimated in small consecutive segments of the (online) data. The segment size we used corresponded to the number of epochs that within a single trial, i.e. the selection of an element in a spelling application. For PASS2D and WARP data sets a single trial consisted of 135 epochs (AMUSE data: 90 epochs). The segment-wise mean and covariance estimates are used to update the weight vector $w_n$, which is now a function of the segment index $n$, according to

$$w_{n+1} = C_{n+1}^{-1}(\mu_2 - \mu_1),$$

with

$$C_{n+1} = (1 - \alpha) C_n + \alpha C_{\text{segment}}.$$}

The update ratio $\alpha$ is a crucial parameter of the adaptive scheme. It is not obvious how to determine the parameter value a priori, because it should be adjusted to the time-scale of the ongoing changes in the feature distribution. We chose a purely data driven approach to tackle this problem: First, we computed the $\alpha$ value that optimizes the classification results on the WARP data set. This was done by running the adaptive classifier with a range of $\alpha$ values for all subjects of the data set and comparing the resulting classification performance to that of the
static classifier. Second, we took the value that yielded the best performance (averaged over the 6 WARP data set subjects) and used it for the adaptive classification of the two remaining data sets.

3 Results

The classification results of the remaining N=31 subjects is shown in Fig. 1 a. In this scatter plot each dot corresponds to one subject. Its x-coordinate represents the classification result of \( \text{LDA}_{\text{fix}} \) and its y-coordinate to the classification result obtained with \( \text{LDA}_{\text{adapt}} \). For points above the diagonal the adaptive classification scheme performs better than the fixed scheme. This was the case for 17 subjects, while for 11 subjects there was no difference in performance and for 3 subjects the performance decreased by using the adaptive classifier. The difference in classification performance was significant with \( p < 0.001 \) in favor of the adaptive scheme.

Figure 1 b to d shows details about the ongoing dynamics during the online run of the subject for which the benefit of adaptive LDA classification was largest. In b we show the evolution of the adapted bias of the first 63 feature dimensions (i.e. the early interval of each channel, cf. section 2.2.1). Scalp maps show the spatial distribution for 4 selected time segments (indicated by the index \( n \)). In subplot d of the same figure the dynamics of the weight vector \( w(n) \) is visualized, again with scalp maps of these time segments. Finally, in subplot c we show the evolution of the confidence of the classifier, computed as the average difference between the classifier output for targets and classifier outputs for the non-target class that is closest to the target output. The larger this difference is, the greater is the class discrimination. Negative values indicate misclassifications. Note the deviation of the two lines towards the end of the BCI session (segments 80 to 128). In this interval, the fixed classifier made 21 mistakes, while the adaptive classifier was wrong in only four times. Furthermore, the dynamics of the adapted means reveal drifts that are strongest for the frontal electrodes. This was observed in addition in a visualization of the evolution of the variances (not shown). A very interesting effect appears after segment number 42. A drastic change in feature means and variance must have taken place, which is eminent in addition by the classifier confidence for these segments. Such changes might have been caused by a sudden move of the subject or other physical influences.

4 Discussion

In this work we have investigated the effect of using a simple and unsupervised adaptation scheme which allows to compensate for slow changes in the first two moments of the pooled feature distribution. The obtained improvement in classification rates indicate the usefulness of the applied approach.

However, care has to be taken when choosing the update parameter. Too fast adaptation rates are prone to overcompensate sudden influences (e.g. trials contaminated by artifacts) and may thereby deteriorate the resulting classification performance notably. We advocate in favor of a conservative choice for the update rates because further simulations have shown that (too) slow update rates are less detrimental (if at all) compared to (too) fast updates (data not shown). The pragmatic approach we chose for finding a suitable value of the update parameter still leaves room for improvement. We conjecture that it would be beneficial to find a subject specific value, rather than taking the same for the entire data set.

An additional benefit of the adaptive method is that it can be applied post-hoc after an experiment in order to investigate the time-scale of potential non-stationarities in the signal. Finding the optimal adaption time scale a posteriori may reveal insights about further improvements of the experimental design (such as block size, when to include breaks, etc.).

In conclusion, we find that it is beneficial to apply adaptive classification methods with carefully chosen update parameters in the context of both, auditory and visual ERP BCIs.
Figure 1: Results of adaptive classification. a: Performance of LDA\text{fix} versus LDA\text{adapt}. b – d: Dynamics of subject VPnz of the PASS2D data set. See text for details.

Acknowledgments

This work is supported by the European ICT Programme Project FP7-224631 and by GRK 1589/1.

References


